DEPARTMENT OF STATISTICS

The Ohio State University

OSU

COLUMBUS, OHIO

(12)

ANALYSIS OF VARIANCE WITH UNEQUAL VARIANCES

by

Edward J. Dudewicz*
~~and~~
Thomas A. Bishop

Technical rept.

14 TR-193

Technical Report No. 193
Department of Statistics
The Ohio State University
Columbus, Ohio 43210
November 1979

12 14

* This author's research was supported by Office of Naval Research Contract No. N00014-78-C-0543

406 331

ANALYSIS OF VARIANCE WITH UNEQUAL VARIANCES

Edward J. Dudewicz

The Ohio State University, Columbus, Ohio 43210

Thomas A. Bishop

Battelle Memorial Institute, Columbus, Ohio 43201

ABSTRACT

Analysis of variance (ANOVA) is often used in quality control studies. It assumes equal variabilities within groups, and no exact procedures have been available for cases with unequal variabilities. In this paper exact procedures are given and illustrated. An indication of the losses to be incurred by using the traditional F-test when variances are unequal is given.

INTRODUCTION

The statistical analysis of variance technique (ANOVA) is often used in various experimental designs in quality control studies.

Dr. Dudewicz is Professor (Department of Statistics) at The Ohio State University, Columbus, Ohio. Dr. Bishop is a research scientist for the Battelle Columbus Laboratories, Columbus, Ohio. A previous version of this paper was presented at the 1977 ATC in Philadelphia and appeared in the ATC Transactions.

This research was supported by Office of Naval Research Contract No. N00014-78-C-0543.

KEY WORDS: ANOVA, Heteroscedasticity, Unequal Variances, Analysis of Variance, Unequal Variability.

For example, it is used in studies where one wishes to determine the significance of the variables under experiment, as in the paper and pulp industries where in bleaching studies the effects of temperature, type of bleaching chemical, pH levels, and consistency are investigated as to their effects on pulp brightness, and in response surface methodology for evaluating the fit of models. This analysis assumes that the observations are normally distributed, and that the variability of results within a treatment is the same for every treatment.

While experimenters are often cautioned that "the assumption of equal variability should be investigated" (e.g., see page 91 of Cochran and Cox [3] or page 46 of Section 27 of Juran, Gryna, and Bingham [7]), no exact statistical procedures have been available for dealing with cases where one finds that variabilities are in fact unequal. While variance-stabilizing transformations and other approximate methods have existed for many years, most experimental situations are such that the problem is far from solved by these approximate methods. For example, such methods misallocate sample size by taking the same sample size from a treatment with relatively small variability, as from a treatment with relatively large variability, even though the need for observations on the latter is substantially greater and they have a greater beneficial effect on performance characteristics of the overall analysis. Also, such methods provide only rough estimates and confidence intervals on the parameters of interest, the parameters of the original problem before a transformation is applied.

In this paper we give exact procedures which we have recently developed for ANOVA when treatment variabilities differ. The procedures are illustrated on typical quality control situations, with explicit attention being given to the level and power of the test.

Recommendations are given as to when one should abandon the common ANOVA procedures in favor of these new ones, with an indication of the costs one may incur by not doing so.

NEW ANOVA PROCEDURES

We will describe the new procedures in the context of the one-way layout; similar procedures are available [2] for higher-way layouts. In the one-way layout, X_{ij} is the j^{th} observation on the i^{th} treatment $(i = 1,2,\ldots,k)$, it is assumed that the X_{ij}'s are independent and normally distributed with mean $E(X_{ij}) = \mu_i$ and variance $Var(X_{ij}) = \sigma_i^2$ where $-\infty < \mu_i < +\infty$ and $0 < \sigma_i^2$, but μ_i and σ_i^2 are otherwise unknown, and the goal (purpose of the experimentation) is to make inferences about $\mu_1,\mu_2,\ldots,\mu_k$, which often represent average process yields. For example, we might want to test the null hypothesis

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k \tag{1}$$

that the treatments do not produce different mean yields. In classical ANOVA procedures it is also assumed that $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$, but we do not make this assumption.

Our procedure for this problem, which we call <u>Procedure</u> $\mathcal{P}_1$, is as follows: Choose a number $z > 0$ (this number is related to the power of the test, and how to choose it will be discussed later), and take an initial sample of size n_0 from each of the k treatments or processes. Any integer sample size $n_0 \geq 2$ will work, but values $n_0 \geq 12$ will give the best results. For the i^{th} process let s_i^2 denote the usual unbiased estimate of σ_i^2 based on the first n_0 observations, and define

$$N_i = \max\left\{n_0 + 1, \left[\frac{s_i^2}{z}\right] + 1\right\} \tag{2}$$

where [x] denotes the greatest integer less than x (e.g. [5.3] = 5). Then take $N_i - n_0$ additional observations from the i^{th} process so we have a total of N_i observations from the i^{th} process; recall that these observations are denoted by $X_{i1}, X_{i2}, \ldots, X_{iN_i}$. Now compute

$$\tilde{\bar{X}}_{i\cdot} = \sum_{j=1}^{n_0} a_i X_{ij} + \sum_{j=n_0+1}^{N_i} b_i X_{ij} \tag{3}$$

where

$$b_i = \frac{1}{N_i}\left\{1 + \sqrt{\frac{n_0(N_i z - s_i^2)}{(N_i - n_0)s_i^2}}\right\} \tag{4}$$

and

$$a_i = \frac{1-(N_i - n_0)b_i}{n_0}\,. \tag{5}$$

Then compute the test statistic

$$\tilde{F} = \sum_{i=1}^{k} (\tilde{\bar{X}}_{i\cdot} - \tilde{\bar{X}}_{\cdot\cdot})^2/z \tag{6}$$

where

$$\tilde{\bar{X}}_{\cdot\cdot} = \frac{\sum_{i=1}^{k} \tilde{\bar{X}}_{i\cdot}}{k}\,, \tag{7}$$

and reject $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$ if and only if

$$\tilde{F} > \tilde{F}(\alpha;k,n_0) \tag{8}$$

where $\tilde{F}(\alpha;k,n_0)$ is the upper $100\alpha^{th}$ percent point of the distribution of

$Q = \sum_{i=1}^{k} (t_i - \bar{t})^2$ when $t_1, \ldots, t_k$ are independent Student's-t random variables with $n_0 - 1$ degrees of freedom and $\tilde{t} = (t_1 + \ldots + t_k)/k$.

We will now discuss the choice of z and tables of $F(\alpha;k,n_0)$. The level and power of the new test do not depend on the unknown variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$, but rather only on $\mu_1, \mu_2, \ldots, \mu_k$, n_0, and z. Thus $z > 0$ should be chosen so that one has the desired power, say P^*, at a given alternative. Exact tables needed for this purpose have been given in [1]. However, as long as n_0 is not very small a simple approximation is available: one may act as if the test statistic $\tilde{F}$ of equation (6) has the same distribution as

$$\frac{n_0 - 1}{n_0 - 3} \chi^2_{k-1}(\Delta) \tag{9}$$

where $\chi^2_{k-1}(\Delta)$ is a noncentral chi-square random variable with k-1 degrees of freedom and noncentrality parameter (using the distributional form given by Johnson and Kotz [6])

$$\Delta = \sum_{i=1}^{k} (\mu_i - \bar{\mu})^2 / z \tag{10}$$

where $\bar{\mu} = (\mu_1 + \ldots + \mu_k)/k$.

A simple method of interpreting Δ is as follows. If the experimenter specifies the minimum range between the largest μ_i and the smallest μ_i which he wishes to detect as δ units, then whenever $\max(\mu_1, \ldots, \mu_k) - \min(\mu_1, \ldots, \mu_k) \geq \delta$ we have $\Delta \geq \delta^2/(4z)$. One can then choose z to attain power P^* when $\Delta = \delta^2/(4z)$, which occurs when $\mu_1 = -\delta/2$, $\mu_2 = \ldots = \mu_{k-1} = 0$, $\mu_k = \delta/2$.

From this point a numerical example, given in the next section, is the easiest way to show very simply how one proceeds, step by step, in practice.

NUMERICAL EXAMPLE IN QC

Suppose we wish to test the hypothesis that 4 different bleaching chemicals are equivalent in their effects on pulp brightness. Suppose we decide to take initial samples of size 10 with each treatment, want only a 5% chance of rejecting H_0 if in fact H_0 is true, and want an 85% chance of rejecting H_0 if the spread among $\mu_1, \mu_2, \mu_3, \mu_4$ is at least 4.0 units. We then proceed, step by step, as follows.

<u>Step 1</u>. (Problem specification.) Here $k = 4$ sources of observations are available, we desire an $\alpha = .05$ level test of $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$, and if the spread among $\mu_1, \mu_2, \mu_3, \mu_4$ is $\delta = 4.0$ units or more we desire power (probability of then rejecting the false hypothesis H_0) of at least $P^* = .85$.

<u>Step 2</u>. (Choice of procedure.) Assuming we do not know that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$, only procedure P_1 given in this paper can guarantee the specifications outlined in Step 1 above. It requires we sample n_0 observations in our first stage, and recommends n_0 be at least 12 (though any $n_0 \geq 2$ will work). Suppose the experimenter only wants to invest 40 units in first-stage experimentation and sets $n_0 = 10$.

<u>Step 3</u>. (First stage.) Draw $n_0 = 10$ independent observations from each source, with results as in Table 1.

Table 1. First Stage Samples of Pulp Brightness

Chemical 1	Chemical 2	Chemical 3	Chemical 4
77.199	80.522	79.417	78.001
74.466	79.306	78.017	78.358
82.746	81.914	81.596	77.544
76.208	80.346	80.802	77.364
82.876	78.385	80.626	77.554
76.224	81.838	79.011	75.911
78.061	82.785	80.549	78.043
76.391	80.900	78.479	78.947
76.155	79.185	81.798	77.146
78.045	80.620	80.923	77.386

Step 4. (Analysis of first stage data.) We now calculate the first stage sample variances $s_1^2, s_2^2, s_3^2, s_4^2$, the total sample sizes needed from the four sources N_1, N_2, N_3, N_4, and the factors $a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4$ to be used in the second stage analysis. The s_i^2's are given in Table 2, along with the other quantities. Here N_i is calculated from (2), b_i from (4), and a_i from (5). The z needed in (2) is found as follows.

We desire power $P^* = .85$ (Step 1 above) when

$$\Delta = \frac{\delta^2}{4z} = \frac{(4.0)^2}{4z} = \frac{4.0}{z} . \tag{11}$$

To set z for this power requirement, we first need to know "When do we reject?". From (8) we know we will later reject H_0 if $\tilde{F} > \tilde{F}(.05;4,10)$ where, approximately,

$$\begin{aligned} \tilde{F}(.05;4,10) &= \frac{n_0-1}{n_0-3}(7.81) \\ &= \frac{10-1}{10-3}(7.81) = 10.04 . \end{aligned} \tag{12}$$

The 7.81 is the value a central chi-square random variable with $k - 1 = 4 - 1 = 3$ degrees of freedom exceeds with probability $\alpha = .05$ (see standard tables, e.g., p. 137 of Pearson and Hartley [8] or p. 459 of Dudewicz [3]).

The power will be, approximately,

$$P[\chi_3^2(\Delta) > 7.81] = .85 \tag{13}$$

if (see p. 53 of the tables in [5])

$$\Delta = 12.301 \text{ ,} \tag{14}$$

so (using equation (11))

$$z = \frac{4.0}{12.30} = .325 \text{ .} \tag{15}$$

Table 2. Analysis of First Stage

	Chemical 1	Chemical 2	Chemical 3	Chemical 4
n_0	10	10	10	10
Sample Mean	77.837	80.580	80.122	77.625
s_i^2	7.9605	1.8811	1.7174	.6762
z	.325	.325	.325	.325
N_i	26	11	11	11
b_i	.046	.364	.390	.686
a_i	.026	.064	.061	.031

<u>Step 5</u>. (Second stage.) Draw $N_i - n_0$ observations from source i $(i = 1,2,3,4)$, yielding Table 3.

Table 3. Second Stage Samples of Pulp Brightness

Chemical 1	Chemical 2	Chemical 3	Chemical 4
82.549	79.990	80.315	78.037
78.970			
78.496			
78.494			
80.971			
80.313			
76.556			
80.115			
78.659			
77.697			
80.590			
79.647			
82.733			
80.522			
79.098			
78.905			

Step 6. (Final calculations.) We now calculate the $\tilde{\bar{\bar{X}}}_{i\cdot}$ of (3) and $\tilde{F}$ of (6), and find

$$\tilde{\bar{\bar{X}}}_{1\cdot} = 78.856,\ \tilde{\bar{\bar{X}}}_{2\cdot} = 80.688,\ \tilde{\bar{\bar{X}}}_{3\cdot} = 80.197,\ \tilde{\bar{\bar{X}}}_{4\cdot} = 77.597 \tag{16}$$

$$\tilde{\bar{X}}_{\cdot\cdot} = 79.335\ , \tag{17}$$

$$\tilde{F} = 17.92\ . \tag{18}$$

Step 7. (Final decision.) Since $\tilde{F} = 17.92$ exceeds $\tilde{F}(.05;4,10) = 10.04$, we reject the null hypothesis and decide the chemicals differ in their effects on pulp brightness.

It should be noted that types of inferences other than tests of hypotheses are available if one uses the new procedures. For example, point estimates, confidence intervals, and selection procedures which guarantee a desired probability of correct selection are not available for the basic parameters of interest if one uses the traditional ANOVA after a transformation of the data, but they are for our Procedure P_1. While we cannot discuss this in detail here, it should be borne in mind that the new methods are backed up by an extensive statistical arsenal of procedures for goals other than testing which one might be interested in.

LOSSES INCURRED BY NOT USING THE NEW PROCEDURE

In our example with $k = 4$ different bleaching chemicals, suppose the new procedure were not used, but rather that the traditional ANOVA procedure were used. If the samples taken were $n_1 = 6$, $n_2 = 60$, $n_3 = 80$, $n_4 = 10$ observations from treatments 1, 2, 3, 4 respectively, the traditional F-test would reject H_0 if its F value exceeded 2.74. However while this yields a level of .05 if $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$, if the variances differ then the level can be greatly different. E.g., if $\sigma_1 = 3$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\sigma_4 = 1$, the true level will be .134 (almost three times the desired .05 level...meaning 13.4% of the time one will decide bleaching chemical has an effect on pulp brightness when in fact it has no such effect). However if $\sigma_1 = 1$, $\sigma_2 = 2$, $\sigma_3 = 2$, $\sigma_4 = 3$, the true level will be .040 (below the desired .05 level) with the traditional F-test.

The F-test has similar problems with its power. For example, while its power at $\Sigma(\mu_i-\bar{\mu})^2 = 1.0$ is .459 when $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 1$, it is .261 when $\sigma_1 = 3$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\sigma_4 = 1$, and it is merely .076 when $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\sigma_4 = 4$. This means one can have no certitude of rejecting H_0 when it is false if one's treatments have unequal variabilities and one uses the F-test.

Since the new procedures yield the desired level and power whether the variances are equal or not, and since sizable losses can be incurred by continuing to use the old procedures when one has unequal variances, use of the new methods is strongly recommended.

REFERENCES

1. Bishop, T. A. and Dudewicz, E. J., "Analysis of variance with unequal variances : test procedures and tables," Technometrics, Vol. 20, 1978, pp. 419-430.
2. Bishop, T. A. and Dudewicz, E. J., "Heteroscedastic ANOVA," 1977, submitted for publication.
3. Cochran, W. G. and Cox, G. M., Experimental Designs (Second Edition), John Wiley & Sons, Inc., New York, 1957.
4. Dudewicz, E. J., Introduction to Statistics and Probability, Holt, Rinehart and Winston, New York, 1976.
5. Haynam, G. E., Govindarajulu, Z. and Leone, F. C., "Tables of the cumulative non-central chi-square distribution," Selected Tables in Mathematical Statistics, Volume I, Markham Publishing Company, Chicago, 1970, pp. 1-78.
6. Johnson, N. L. and Kotz, S., Distributions in Statistics: Continuous Univariate Distributions - 2, Houghton Mifflin Company, Boston, 1970.

7. Juran, J. M., Gryna, F. M., Jr., and Bingham, R. S., Jr. (Editors), Quality Control Handbook (Third Edition), McGraw-Hill, Inc., New York, 1974.

8. Pearson, E. S. and Hartley, H. O. (Editors), Biometrika Tables for Statisticians, Volume I (Third Edition, Reprinted with additions), Cambridge University Press, London, 1970.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 193
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): ANALYSIS OF VARIANCE WITH UNEQUAL VARIANCES
5. TYPE OF REPORT & PERIOD COVERED: Technical Report
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Edward J. Dudewicz, Thomas A. Bishop
8. CONTRACT OR GRANT NUMBER(s): N00014-78-C-0543
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, The Ohio State University, Columbus, Ohio 43210
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 042-403
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Department of the Navy, Arlington, Virginia 22217
12. REPORT DATE: November 1979
13. NUMBER OF PAGES: 12
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): Approval for public release; distribution unlimited.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
18. SUPPLEMENTARY NOTES:
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): ANOVA, Heteroscedasticity, Unequal Variances, Analysis of Variance, Unequal Variability
20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

Analysis of variance (ANOVA) is often used in quality control studies. It assumes equal variabilities within groups, and no exact procedures have been available for cases with unequal variabilities. In this paper exact procedures are given and illustrated. An indication of the losses to be incurred by using the traditional F-test when variances are unequal is given.